true 3D display). At block **1404**, process **1401** can interface with a display system (e.g., a screen, various types of projectors such as LED, microLED, LASER, etc.) to display the images. Outputting the images can be synchronized with outputting audio according to time tags added during the capture stage.

[0116] FIGS. **15**A and **15**B are conceptual diagrams illustrating examples **1500** and **1550** of a 3D conversation in an artificial reality environment. Example **1500** illustrates a first side of a 3D conversation where a sending/receiving device **1504** includes capture devices **1506** (color camera, depth camera, and microphone). The cameras of capture devices **1506** are each associated with calibration data defining the camera's intrinsic parameters (the optical, geometric, and digital characteristics of the camera) determined during manufacture of the camera and extrinsic parameters (location and orientation in the 3D environment). The capture devices **1506** capture color images, depth images, and an audio feed of user **1502**, which are tagged with capture time and which device captured each part of the captured data. Device **1504** then performs filtering and tagging to remove portions from images not depicting the user, remove background noise from the audio stream, and, based on the device tags and the associations between the calibration data and device identifiers, tags the calibration data for the device that captured each part of the data to the corresponding captured data. Device **1504** then compresses each of the filtered and tagged data streams and sends them to device **1554** (FIG. **15**B).

[0117] Meanwhile, device **1504** is also receiving compressed data streams from device **1554** (FIG. **15**B). Device **1504** decompresses these data streams into color images, depth data, and audio data (with associated calibration data). Device **1504** next reconstructs the depth data and calibration data into a 3D representation (in this case a point cloud). Device **1504** takes an indication of the viewpoints of each eye of user **1502**, as detected by artificial reality device **1508**, to place virtual cameras in relation to the point cloud to generate two 2D images of user **1552** (FIG. **15**B) from a viewpoint of the user **1502**. Device **1504** also adds color data to these images based on the calibration data and synchronizes them with the audio data based on time tags associated with the data feeds. In examples **1500** and **1550**, rendering further includes using machine learning object recognition to remove, from the representations of the users **1502** and **1552**, the artificial reality devices **1508** and **1558** and further using predicative machine learning to fill in the missing portions of the representations of the users, allowing the users to appear as if they were not wearing the artificial reality devices. Device **1504** finally provides these images and synchronized audio to artificial reality device **1508** so artificial reality device **1508** can project a representation **1510** of user **1552** (FIG. **15**B). In example **1550** (FIG. **15**B), user **1552** is holding capture devices **1556** close to his body, allowing only the capture of user **1552**'s head and upper torso. Thus, the generated 3D representation, subsequent 2D images, and ultimately the projection **1510** only show the upper part of the user **1552**.

[0118] Example **1550** illustrates a second side of the 3D conversation which performs a similar process to example **1500**. In particular, sending/receiving device **1554** includes hand-held capture devices **1556** (color camera, depth camera, and microphone). The cameras of capture devices **1556** are each associated with calibration data defining the camera's intrinsic parameters (the optical, geometric, and digital characteristics of the camera) determined during manufacture of the camera and extrinsic parameters (location and orientation in the 3D environment). The capture devices **1556** capture color images, depth images, and an audio feed of user **1552**, which are tagged with capture time and which device captured each part of the captured data. Device **1554** then performs filtering and tagging to remove portions from images not depicting the user **1552**, remove background noise from the audio stream, and, based on the device tags and the associations between the calibration data and device identifiers, tags the calibration data for the device that captured each part of the data to the corresponding captured data. Device **1554** then compresses each of the filtered and tagged data streams and sends them to device **1504** (FIG. **15**A).

[0119] Meanwhile, device **1554** is also receiving the compressed data streams from device **1504** (FIG. **15**A). Device **1554** decompresses these data streams into color images, depth data, and audio data (with associated calibration data). Device **1554** next reconstructs the depth data and calibration data into a 3D representation (in this case a 3D mesh). Device **1554** takes an indication of a viewpoint of user **1552**, as detected by artificial reality device **1558**, to place a virtual camera in relation to the 3D mesh to generate a 2D image of user **1502** (FIG. **15**A) from a viewpoint of the user **1552**. Device **1554** also adds color data to this image based on the calibration data and synchronizes the image with the audio data based on time tags associated with the data feeds. Device **1554** removes, from the representation of the users **1502**, the artificial reality devices **1508**, allowing the user **1502** to appear as if she were not wearing the artificial reality device **1508**. Device **1552** finally provides these images and synchronized audio to artificial reality device **1558** so artificial reality device **1558** can project a representation **1560** of user **1502**. In example **1500** (FIG. **15**A), user **1502** has placed capture devices **1506** on a surface far enough from her body to capture images of her entire body. Thus the generated 3D representation, subsequent 2D images, and ultimately the projection **1560** shows a complete representation of the user **1502**. Further, in example **1550**, user **1552** has moved around the projection of user **1560** during the 3D conversation. Thus, the viewpoint of user **1552** is toward the side of the projection **1560**. Accordingly, during rendering, the virtual camera is placed to the side of the 3D representation, producing images shown projection **1560** being from the side of the user **1502**.

[0120] Reference in this specification to "implementations" (e.g., "some implementations," "various implementations," "one implementation," "an implementation," etc.) means that a particular feature, structure, or characteristic described in connection with the implementation is included in at least one implementation of the disclosure. The appearances of these phrases in various places in the specification are not necessarily all referring to the same implementation, nor are separate or alternative implementations mutually exclusive of other implementations. Moreover, various features are described which may be exhibited by some implementations and not by others. Similarly, various requirements are described which may be requirements for some implementations but not for other implementations.

[0121] As used herein, being above a threshold means that a value for an item under comparison is above a specified other value, that an item under comparison is among a